Geometry *versus* Topology: Testing Self-Consistency of the NiceGraph Program

Tomaž Pisanski,[a] *Marko Razinger,*[*,b] *and Ante Graovac*[c]

[a]*Department of Theoretical Computer Science, IMFM, University of Ljubljana, Ljubljana, Slovenia*

[b]*National Institute of Chemistry, Ljubljana, Slovenia*

[c]*The Rugjer Bošković Institute, HR-10001 Zagreb, POB 1016, Croatia*

Algorithm for drawing graphs determine a collection of points in space. On the other hand, each collection of points in space may define an abstract graph *via* the so-called overlapping sphere model. The process may be iterated. The problem of which graphs are stable in the above mentioned process is considered.

Molecular structure carries a great amount of information: from simple parameters, such as the number and nature of atoms and bonds, to the more elusive notions of conformation and shape.[1] An ideal model used to describe molecular structure should be self-contained and should identify the structure in question uniquely. Furthermore, it should be as simple as possible and thus irredundant. The quest for such a model remains one of the foremost goals of chemical information science. The endeavour to describe molecular structure with graphs fits into the above scheme although it is well-known that additional information is needed in most cases and that there are structural features not suitable for graph-theoretical description.

Graphs are abstract mathematical objects suitable for describing the molecules, namely the connectivity of their constituent atoms. All geometrical information is missing in such a description. However, a part of molecular geometry could be recovered. An example is offered by Hückel Molecular Orbital (HMO) model where the HMO bond orders are able to give some in-

---

* Dr. Marko Razinger passed away on July 31, 1996.

formation on the bond lengths in conjugated molecules. It is also clear that geometrical symmetry is already contained in the automorphism group of the graph.[2,3,4]

The question of a relationship between topology and geometry of molecules could be reversed. In the present paper, we describe a method which is able to reconstruct molecular connectivity from the molecular geometry data. It is based on the overlapping spheres (OS) model.

The OS model is applied here to several benzenoid and non-benzenoid condensed polycyclic hydrocarbons, linear and branched alkanes, and fullerenes[5,6] in order to test the self-consistency of the recently introduced Nice-Grah program[7] which is a part of the computer program Vega.[8] The testing follows the line: topology $\rightarrow$ geometry $\rightarrow$ topology. For a given molecule M, its graph $G(M)$ is given first and afterwards the positions of vertices $R = R(G) = (r_1, r_2, ..., r_n)$ in 3D-space are computed by some method, for instance, by the NiceGraph program. Then, the OS method is invoked to R in order to compute a plausible molecular graph $G' = G'(R)$. If the initial graph G is equal to the final one, G', the coordinate rendering method is said to be *self-consistent*, otherwise it is not self-consistent.

OVERLAPPING SPHERE METHOD

Given a set of n points R in 3D-space and a positive number $\rho > 0$, we may draw n closed balls $B_1 := B(r_i, \rho)$, $i = 1, 2, ..., n$, each ball B_i centered at r_i, and having a radius ρ. Define a graph $G(R, \rho)$ as follows: The vertices are the n selected points. Two vertices r_i and r_j are adjacent if and only if the corresponding balls intersect, *i.e.* if $B_i \cap B_j \neq 0$. The radius ρ will be called a *unit* and the graph the *overlapping spheres graph*.

Clearly, the same method can be used in 2D or even 1D, or in any other metric space.

One can keep the set of vertices R and change the unit ρ. For small values of ρ, we get the empty graph $E_n = nK_1$ and for large values of ρ we get the complete graph K_n. Let $D(R)$ denote the matrix with elements $D_{i,j} = \|r_i - r_j\|/2$. Let ρ be the number of distinct elements of matrix D and let $0 = \rho_0 < \rho_1 < \rho_2 < ... < \rho_p < \rho_{p+1} = \infty$ denote the increasing sequence of matrix elements. It is obvious that $G(R, \rho)$ does not change for $\rho_k \leq \rho < \rho_{k+1}$. This means that a sequence of graphs with an increasing number of edges:

$$nK_1 = H_0 \subset H_1 \subset H_2 \subset ... \subset H_p = K_n \tag{1}$$

is defined, where $H_k := G(R, r_k)$. Note that the $p + 1$ graphs H_k cover all cases of $G(R, \rho)$ for all possible radii $0 \leq \rho < \infty$.

The problem is how to choose a suitable value of ρ in order to obtain a molecular graph. If graph G, whose vertices are placed in 3D (or 2D) in order to define R, coincides with one of graphs H_k in the above sequence, was say that G is *consistently placed*.

There is an easy consistency test for a molecular graph G with m edges. If G is consistently placed, then it must be identical to one of graphs H_k from above. If we select k such that H_{k-1} has less than m edges and H_k has at least m edges, then G is consistently placed if and only if it is identical to H_k. Note that ρ_k can be simply determined from matrix D by first sorting the non-zero elements of D and selecting the m- th one. There are known algorithms that can find the m-th element in an array in linear time (see, for instance Ref. 9). Also there is no need for complete isomorphism testing for G and H_k since both graphs have the same set of vertices. Since they both have the same vertex labeling, one only has to check if each edge $v_i - v_j$ of graph G is also present in H_k and vice versa. This means that the consistency can be checked in linear time $O(m)$. Further savings in running time are possible if we work with squares of distances (radii) instead of distances, since all decisions are based on the order of radii. The order does not change if squares are considered. In this way, we avoid calculating square-roots.

CONSISTENCY CHECK OF SOME MOLECULAR GRAPHS

Several families of molecular graphs have been tested. For each graph, we first calculated the coordinates using the NiceGraph algorithm. The coordinates that we obtained were not very different from the ones that could be calculated by some quantum-chemical program for geometry optimization, in the sense that the molecule could be easily recognized. Major differences can be attributed to the fact that the NiceGraph program ignores all angles during the calculation process. The coordinates and the number of edges were then fed to the OS algorithm in order to perform the consistency check.

Out of 40 examples, 30 passed the consistency test. The results can be seen on the figures below.

For simplicity, we use idealized coordinates in most figures. In order to see the difference, we show computed coordinates of trinaphthylene and the proof of its inconsistency in Figure 1. For the fullerenes, we use Schlegel diagrams that were computed by the package Vega. Although our procedure can test consistency automatically, it is interesting to note that the counter-example for G need not to be H_k, as described above, but some earlier member H_i, $i < k$, of the sequence (1). Namely, we can look for the smallest graph H_i that has an edge e_1 not belonging to G and does not have an edge e_2 which is present in G.

TABLE I

Consistency check for some classes of molecular graphs

Molecule	consistent	inconsistent	total
trees-unbranched	5	0	5
trees-branched	3	1	4
non-benzenoid condensed polycyclic	8	1	9
benzenoid-catacondensed-linear	4	1	5
benzenoid-catacondensed-nonlinear-unbranched	1	2	3
benzenoid-catacondensed-nonlinear-branched	1	2	3
benzenoid-pericondensed	6	2	8
fullerenes	2	1	3

It seems that the smallest inconsistent example is given by 2,3,4-trimethylpentane shown in Figure 2. The failure may be explained by the fact that angles between the incident edges are not considered during the calculation of the coordinates by the NiceGraph program. In both pairs of terminal methyl groups, the carbons are placed too close together while the inner edges are stretched too much. Consequently, the OS method produces false triangles and an unconnected graph.
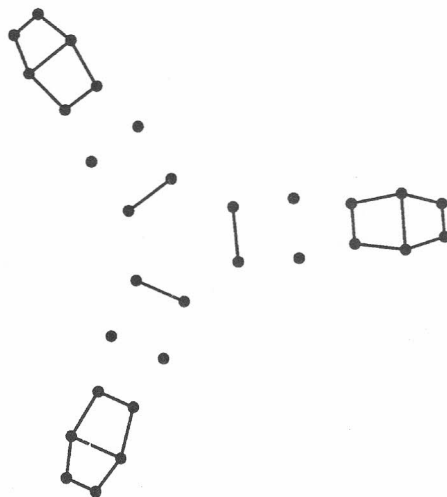
Figure 1. Proof of the inconsistency of the Overlapping Spheres approach in the trinaphthylene example. The coordinates were calculated by the NiceGraph program from connectivity data alone. The graph, reconstructed by the OS method, has 3 superfluous edges and 15 missing edges relative to the original graph.

Trees

paths: 4 - 20

CONSISTENT   n-butane   n-pentane   n-hexane   n-decane

n-eicosane

branched trees

CONSISTENT   2-methyl pentane   3-methyl pentane   3-ethyl pentane
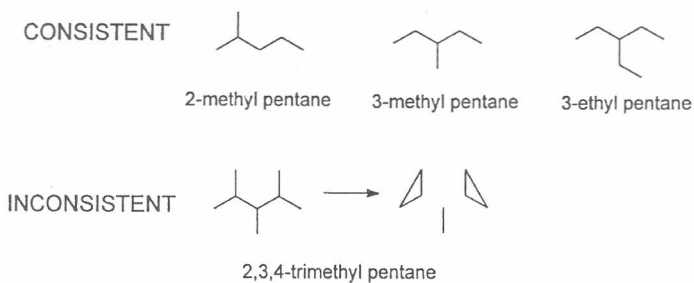
INCONSISTENT

2,3,4-trimethyl pentane
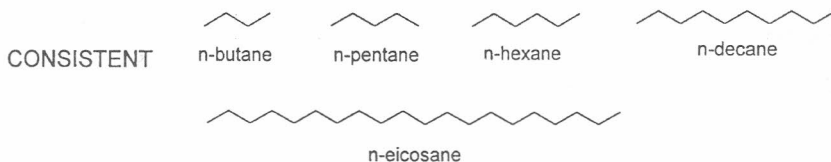
Figure 2. Tree graphs – examples of normal and branched alkanes. In Figs. 2, 3, and 4, the molecular graphs are shown as idealized structural formulae for clarity sake. In the cases where consistency could not be achieved, the counter-examples are shown along with the original graphs, see for example the inconsistency of 2,3,4-trimethyl-pentane.

While even long unbranched trees pass the consistency test, linear cata-condensed benzenoids fail the test when the number of condensed rings is greater than four. The breakdown occurs near the centre of pentacene; see Figure 3. The central rings are stretched in the direction of the long axis of the molecule, bringing bivalent vertices of the hexagon closer together. The symmetry is broken by the NiceGraph program since the algorithm only finds an approximate solution to the minimum.

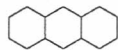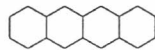# Benzenoids

linear catacondensed benzenoids
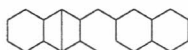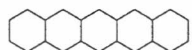
CONSISTENT
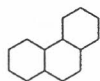
benzene   naphthalene   anthracene   naphthacene

INCONSISTENT

pentacene

non-linear catacondensed benzenoids

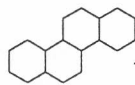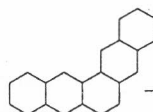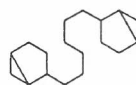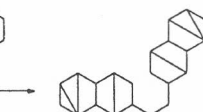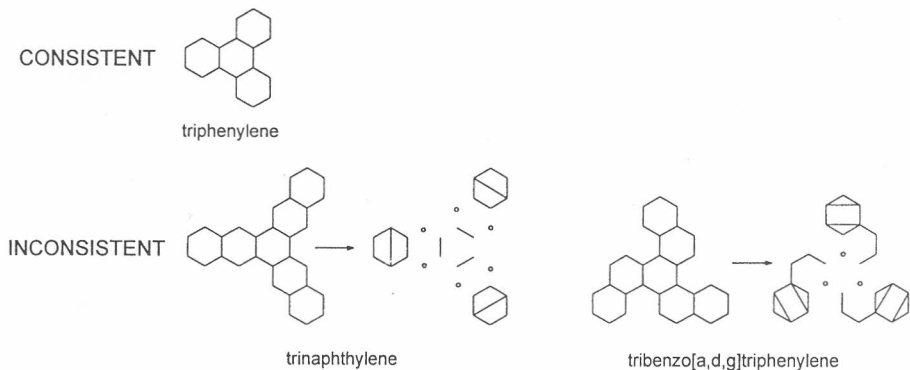## a. unbranched

CONSISTENT

phenanthrene

INCONSISTENT

chrysene   pentaphene

b. branched

CONSISTENT

triphenylene

INCONSISTENT

trinaphthylene

tribenzo[a,d,g]triphenylene

pericondensed benzenoids

CONSISTENT

pyrene   perylene   anthrathene   coronene

CONSISTENT

benzo[c]pyrene   ovalene

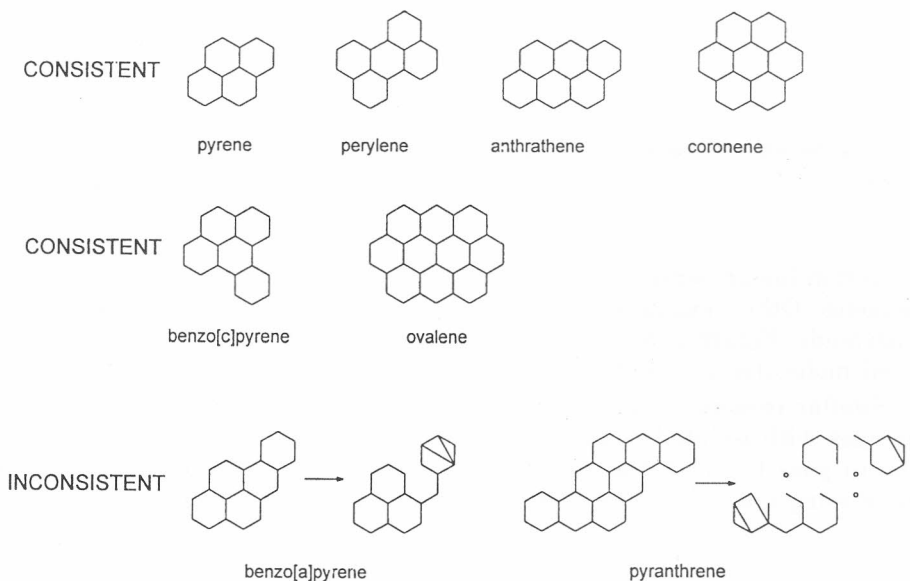INCONSISTENT

benzo[a]pyrene   pyranthrene

Figure 3. Benzenoids – examples of linear catacondensed, non-linear catacondensed, unbranched and branched, and pericondensed benzenoid structures. In inconsistent examples, the erroneously resonstructed graphs are shown too.

Non-benzenoid

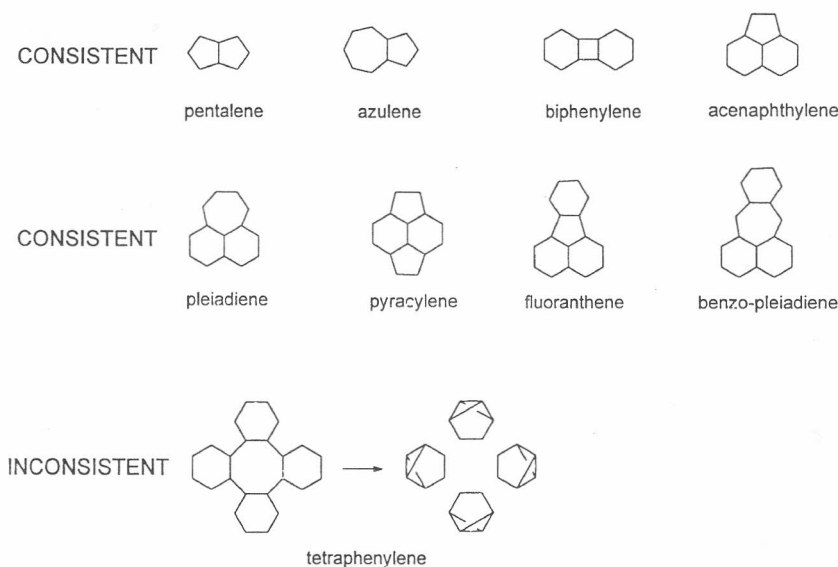## condensed polycyclic hydrocarbons

Figure 4. Non-benzenoid condensed polycyclic structures, consistent and inconsistent examples.

In non-linear benzenoids, the failure occurs even earlier in the four-ring chrysene. Other examples of branched catacondensed and pericondensed benzenoids (Figure 3) and non-benzenoids (Figure 4) indicate that more balanced molecular graphs have more chance to pass the consistency check.

Similar reasoning applies also to fullerenes; see Figure 5. It seems that the ones with isolated pentagons tend to be consistent. When the fullerene fails to pass the test, the superfluous edges appear within pentagons and the missing ones are those originally shared by two hexagons.

CONCLUSION

The topology of molecular graphs has some intrinsic geometric information which is usually revealed by molecular mechanics methods or by sim-

C60

C70, P = 0, Q = 0
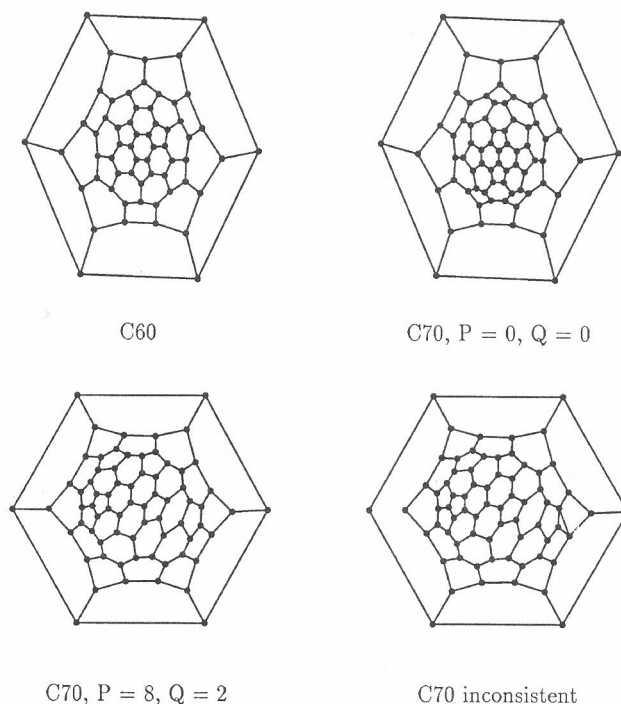
C70, P = 8, Q = 2

C70 inconsistent

Figure 5. Fullerenes – C60 and C70 systems. While C60 and C70(P=0, Q=0) are shown to be consistent in the OS approach, C70(P=8, Q=2) is inconsistent and the counterexample with one missing and one superfluous edge is shown.

plified algorithms like the NiceGraph. In many cases, the coordinates of points in space obtained by such algorithms can be used to reconstruct graph connectivity.

We have tested several classes of molecular graphs for self- consistency. The majority of graphs passed the test. It turns out that smaller graphs tend to be self-consistent although some large self- consistent graphs were found, *e.g.* some fullerenes. Also, graphs from a given family that are more »balanced« passed the test where »unbalanced« failed. It would be interesting to explore which graph theoretical invariant correlates best with the intuitive notion of »being balanced«, which is related but not identical to the presence of a centre of symmetry. The reasons for failures of self- consistency could be better understood if a much larger set of molecular graphs were tested. However, we believe that the method admits several improvements, which can be tested against our 40 graph benchmark set. The winning strategy can be later tested with a significantly larger testing set.

Clearly, the crude approximation to molecular coordinates provided by the NiceGraph program could be improved by including chemical knowledge

into the purely graph theoretical routine. On the other hand, the OS method could be improved instead. Note that the OS method is a purely geometric approach without any chemical expertise. There are numerous ways of improvement available. For instance, one could use spheres of different radii, corresponding to different atoms. Also, possibilities leading to chemically unrealistic reconstructed graphs: strained angles, illegal valencies, *etc.*, could be automatically dismissed. We believe that such *a posteriori* improvements are even superior to the endeavour to calculate coordinates with high precision. As we know, real molecules maintain their connectivity even when their coordinates are significantly perturbed due to vibrations. Recently, we have already implemented some first-order improvements of the OS method in the Vega package. The results of self-consistency of the improved OS method will be reported elsewhere.

REFERENCES

1. M. Randić and M. Razinger, *J. Chem. Inf. Comput. Sci.* **35**(1995) 594-606.
2. T. Pisanski, Recognizing Symmetric Graphs, presented at Graph Drawing '93. Sèvres, France, September 26-29, 1993.
3. M. Razinger, K. Balasubramanian, M. Perdih, and M. E. Munk, *J. Chem. Inf. Comput. Sci.* **33** (1993) 812-825.
4. D. E. Manolopulos and D. E. Fowler, *J. Chem. Phys.* **96** (1992) 7603-7614.
5. P. W. Fowler and D. E. Manolopoulos, An Atlas of Fullerenes, Clarendon Press, Oxford, 1995.
6. B. Plestenjak,. T. Pisanski, and A. Graovac, *J. Chem. Inf. Comput. Sci.* 36 (1996) 825–828.
7. T. Pisanski, B. Plestenjak, and A. Graovac, *Croat. Chem. Acta* **68** (1995) 283-292.
8. Vega Version 0.2; Quick Reference Manual and Vega Graph Gallery (T. Pisanski, (ed.), IMFM, Ljubljana, 1995. http://www.mat.uni-lj.si/dwnld.htm
9. E. Horowitz and S. Sahni, Fundamentals of Computer Algorithms, Computer Science Press, Potomac MD, 1978.

SAŽETAK

Geometrija i topologija: samousaglašenost programa NiceGraph

Tomaž Pisanski, Marko Razinger i Ante Graovac

Algoritmima za crtanje grafova utvrđuje se skup točaka u prostoru. No, svaki skup točaka u prostoru opet definira neki apstraktni graf, pomoću tzv. modela prekrivajućih kugli. Ovaj se proces može iterirati. U ovome radu se razmatra koji grafovi su stabilni u odnosu na taj proces.